



TIL BIRLIKLARINI CHASTOTALI TAHLIL QILISHDA TIL KORPUSLARINING AHAMIYATI

Maxammadiyeva Diyora Raxmatovna

*Alisher Navoiy nomidagi Toshkent davlat o‘zbek tili
va adabiyoti universiteti doktoranti
diyoramaxammadiyeva@gmail.com
Tel: +998995938892*

Annotatsiya. Ushbu maqolada o‘zbek tilining kelgusi taraqqiyotida muhim omillardan biri bo‘lgan til korpusi tushunchasi hamda til birliklarini korpus yordamida chastotali tahlil etishning afzalliklari haqida so‘z yuritiladi. Til korpuslari va ularga qo‘yiladigan talablar, chastotali tahlil turlari va metodlari haqida mulohazalar bayon etilgan. Xususan, chastotali tahlil qilishning til strukturasini o‘rganish, Tilning davriy rivojlanishini tahlil etish, Til birliklari o‘rtasidagi semantik munosabatlarni aniqlashdagi afzalliklari va bunda til korpuslarining katta ma’lumotlar ombori sifatidagi xizmat e’tirof etilgan. Shuningdek, chastotali tahlillar orqali tilimiz fonetik, leksik, semantik, grammatik strukturasidagi taraqqiyot jarayonlarini tez va oson tadqiq etish mumkinligi to‘g‘risidagi fikrlar keltirilgan.

Kalit so‘zlar. Statistik tahlil, matn va til birliklari o‘rtasidagi struktura, dinamika, semantik maydon, mualliflik korpusi, morfologik birliklar lingvostatistikasi.

ЗНАЧЕНИЕ ЯЗЫКОВЫХ КОРПУСОВ В ЧАСТОТНОМ АНАЛИЗЕ ЯЗЫКОВЫХ ЕДИНИЦ

Абстрактный. В данной статье говорится о понятии языкового корпуса, которое является одним из важных факторов будущего развития узбекского языка, и преимуществах частотного анализа языковых единиц с использованием корпуса. Представлены комментарии к языковым корпусам и требованиям к ним, видам и методам частотного анализа. В частности, признаются преимущества частотного анализа при изучении структуры языка, анализе периодического развития языка, определении семантических отношений между языковыми единицами, а также обслуживании языковых корпусов как большого хранилища данных. Также с помощью частотного анализа можно быстро и легко изучить процессы развития фонетического, лексического, семантического и грамматического строя нашего языка.

Ключевые слова. Статистический анализ, структура, динамика между текстовыми и языковыми единицами, семантическое поле, авторский корпус, лингвостатистика морфологических единиц.

THE IMPORTANCE OF LANGUAGE CORPORATIONS IN FREQUENCY ANALYSIS OF LANGUAGE UNITS

Abstract. This article discusses the concept of a language corpus, which is one of the important factors in the future development of the Uzbek language, and the advantages of frequency analysis of language units using a corpus. Considerations are given to language corpora and the requirements placed on them, types and methods of frequency analysis. In particular, the advantages of frequency analysis in studying the structure of the language, analyzing the periodic development of the language, identifying semantic relationships between language units, and the role of language corpora as a large repository of information in this regard are recognized. Also, the ideas about the possibility of quickly and easily studying the development processes in the phonetic, lexical, semantic, and grammatical structure of our language through frequency analysis are presented.

Keywords. Statistical analysis, structure between text and language units, dynamics, semantic field, authorial corpus, linguostatistics of morphological units.

KIRISH

Raqamli texnologiyalar rivojlangani sari ulardan foydalanish kundalik ehtiyojga aylanib bormoqda. Yangi O‘zbekistonda bugun har bir soha, raqamlashtirilish jarayonida. Raqamli iqtisodiyot, elektron hukumat, elektron murojaat kabi tushunchalar hayotimizning bir bo‘lagiga aylangan. Tilshunoslik ham bundan mustasno emas. Zamonaviy axborot texnologiyalari tilning funksional imkoniyatlaridan foydalanish borasida benihoya keng qulayliklar eshigini ochdi. Kompyuter tarjimasi, tahriri, tahlili, elektron lug‘atlar va tezauruslar fikrimiz dalilidir. Ayniqsa, zamonaviy elektron lug‘atlar tuzish va ulardan foydalanish madaniyatini shakllantirish til imkoniyatini kengaytirishda muhim ahamiyatga ega. Respublikamizda har bir sohada axborot-texnologiyalaridan unumli foydalanishga, dasturiy ta‘minot asosida tuziladigan loyihalarga katta e‘tibor qaratilmoqda. Xususan, Matnlar korpusiga oid ilmiy tadqiqotlar salmog‘ining ko‘payishi natijasida tilshunoslikda korpus lingvistikasi yo‘nalishi shakllandi. Bugungi kunda o‘zbek tilshunosligida aynan shu sohaning rivojlanishi kelgusida til birliklarining xususiyatlarini aniqlashda, tabiiy tildagi davriy hodisalarni o‘rganishda elektron shaklda saqlanadigan yozma yoki og‘zaki, kompyuterlashtirilgan qidiruv tizimiga dasturiy ta‘minot asosida joylashtirilgan matnlar jamlanmasi orqali tez va ishonchli statistik tahlil qilish imkonini beradi. Ayniqsa, Korpus yordamida Chastotali tahlillarni amalga oshirish bir necha marotaba qulaydir. Turli janr va uslub matnlaridagi til birliklarining (so‘zlar, frazeologik birliklar, morfemalar va boshqalar) chastotasini o‘rganishda katta hajmli korpuslarning asqotishi tabiiy. Har qanday tadqiqotda statistik tahlil muhim ahamiyat kasb etadi. Aniq raqamlarga tayanilgan

xulosalar ilmiy nazariyani nafaqat tasdiqlashga, balki bilimlarni mustahkamlash va ularni isbotlashga ham yordam beradi.

ADABIYOTLAR TAHLILI:

Jahon adabiyotshunosligi va korpus lingvistikasi tajribasida korpus tuzish va korpusga tayangan tadqiqotlar ko‘p. Til birliklarini lingvostatistik tadqiq etishga qaratilgan ishlar sirasida S.M.Muhammedov, A.Babanarov, N.K.Ayymbetov, S.Karimov, A.Qarshiyev, G.Isroilova, S.Rizayevlarning ishlari e‘tiborga molik[1,2,3,4,5,6]. Ushbu olimlarning ishlari asosan badiiy, publisistik matnlarning lingvostatistik tahlili doirasida olib borilgan. Sh.Hamroyeva, N.Abdurahmonova, G.Toirova, D.Axmedova, R.Karimov va boshqalarning o‘zbek tili korpusini tuzish tamoyillari, metodikasi va lingvistik asoslariga bag‘ishlangan ishlarida til korpusining turli aspektlari tahlilga tortilgan[7,8,9,10]. Kelgusi tadqiqotlarda she‘riy asarlar korpusini tuzish va unga asoslangan lingvostatistik tadqiq olib borish kun tartibiga qo‘yganmiz. Korpusning bunday turi poetik uslubdagi matnlar tilining til birliklarini o‘rganish, leksik qatlamlarining tarkibini aniqlashtirish va tasniflash, matnning lingvostatistik tahlilini amalga oshirish, muayyan lahja va shevaga xos bo‘lgan lug‘aviy birliklarini aniqlash, tildagi uslublar leksikasi o‘rtasidagi umumiylik hamda xususiylik, o‘ziga xoslik tomonlarini kuzatish, ijodkorlarning so‘z qo‘llash mahoratini belgilash, matnlarni kompyuterda avtomatik tekshirish, istalgan matnga avtomatik ishlov berish orqali leksikografik manbalarini yaratish masalasi o‘rganilishida muhim manba bo‘la oladi.

TADQIQOT METODOLOGIYASI.

Maqolada badiiy matnlarni lingvostatistik tadqiq etishning ahamiyati, yutuqli tomonlari muhokama etiladi. Chastotali tahlil etishning matn va til birliklari o‘rtasidagi struktura, dinamika, semantik maydon uzviylikini aniqlashdagi aniqlik va tezkorlik me‘zonlari tahlil qilinadi. Badiiy matnni statistik tahlil qilishning ahamiyati o‘rganildi.

TAHLILLAR VA NATIJALAR.

O‘zbek tilshunosligida keyingi yillarda til korpuslarini tuzish metodikasi va texnikasi tadqiqiga bag‘ishlangan tadqiqotlar ko‘paydi. Ular asosida bir necha korpuslar ishlab chiqildi. Kompyuterda amalga oshiriladigan statistik tahlilning yutug‘i shundan iboratki, bir muallifning hayoti va ijodining turli davrlarida yaratilgan asarlari uslubini o‘rganishda yordam beradi. Matnni tadqiq qilishda statistik metodlardan foydalanilar ekan, bunday metodlar muayyan bir matnda so‘z yoki ibora, bo‘g‘in yoki gap singari til birliklarining qo‘llanishida o‘ziga xos qonuniyatlarni aniqlashga qaratiladi[1]. Matnlarning Chastotali tahlili tilshunoslar uchun tilning tuzilishini, davriy rivojlanishini va til birliklari o‘rtasidagi o‘zaro aloqalarini tahlil qilish uchun zarur vosita hisoblanadi. Tilshunoslikda chastotali tahlildan olingan natijalardan quyidagi maqsadlarda foydalanish mumkin:

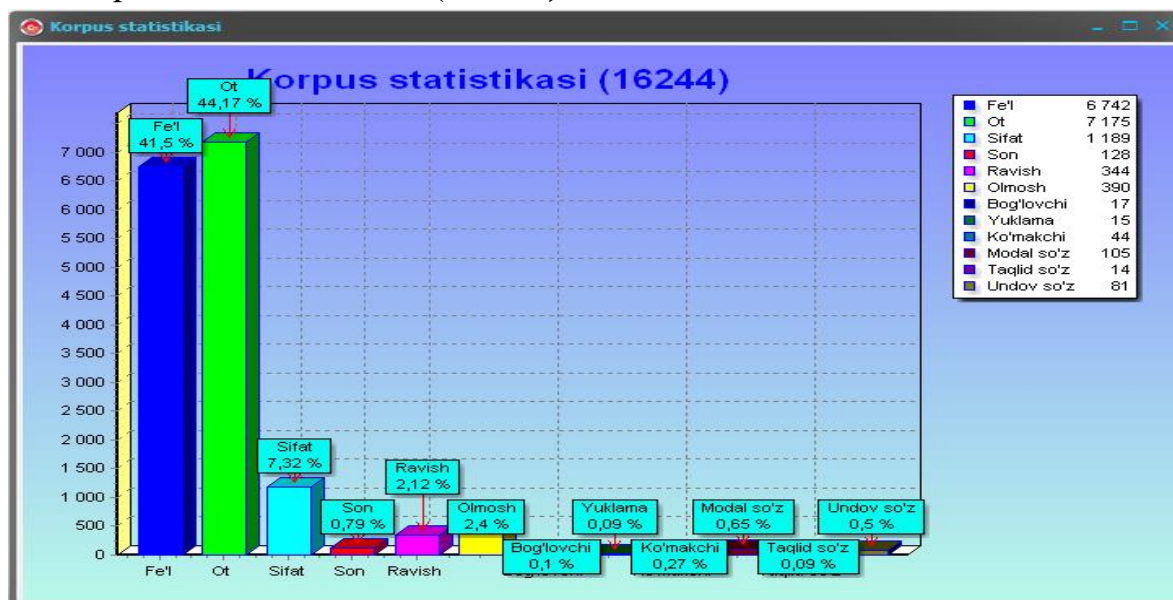
1. Til strukturasini o‘rganishda chastotali tahlil statistikasi tilning sintaktik va morfologik tuzilmasini o‘rganish uchun muhim ahamiyatga ega. Eng ko‘p uchraydigan so‘zlar, frazeologik birliklar va ularning tartibi tilning umumiy strukturasini haqida qimmatli ma’lumotlarni taqdim etadi. Til birliklarining chastotasi yordamida sintaktik va morfologik birliklardagi jarayonlarni

o‘rganishda qo‘l keladi. Masalan, o‘zbek tilida fe‘l, ot va olmoshlarning tez-tez ishlatilishini tahlil qilish orqali sintaktik butunliklar haqida quyidagicha xulosa chiqarish mumkin. Turli janrdagi asarlarda gap bo‘laklarining har xil nisbati qo‘llaniladi: qayerda hodisa ko‘p bo‘lsa, fe‘l ham ko‘p bo‘ladi; tavsiflar ko‘p bo‘lgan joyda otlar ko‘proq bo‘ladi; odamlar o‘rtasida ko‘proq dialog bo‘lganda ko‘proq olmoshlardan foydalaniladi.

2. Tilning davriy rivojlanishini tahlil etishda chastotali tahlillar muhim xulosalarni dalillaydi. Vaqtlar o‘tishi bilan tilning leksik, semantik, grammatik, fonetik fondida qanday o‘zgarishlar bo‘lganligini ko‘rsatadi. So‘zlar va iboralarning chastotasidagi o‘zgarishlar yangi tushunchalarning paydo bo‘lishi yoki mavjud tushunchalarning o‘zgarishini anglatadi. Bu tilshunoslarga tilning tarixiy rivojlanish jarayonini o‘rganishda yordam beradi.

3. Til birliklari o‘rtasidagi semantik munosabatlarni aniqlashda, asosiy mazmunni haqida semantik tahliliy xulosalar chiqarishda chastotali tahlil materiallari ishonchli manba bo‘la oladi. Masalan, biror so‘zning tez-tez ishlatilishi, uning matnda nechog‘li muhimligi haqida axborot beradi hamda ijtimoiy-siyosiy hayot va zamona ahlining turmush-tarzi, ruhiyati, kayfiyati, qiziqishlari haqidagi ma‘lumotlarni to‘plashda kalit vazifasini bajaradi.

Biz Usmon Azim dramalari uchun yaratilgan kichik mualliflik korpusi orqali ayrim morfologik birliklarining lingvostatistik tadqiqini amalga oshirdik. Tadqiqotda Usmon Azimning “Adibning umri” dramatik asarlar to‘plamidagi 7 ta dramasi lingvostatistik tahlil qilindi. Usmon Azim 7 ta dramada jami 16244 ta so‘zshakldan foydalangan. Shundan 6742 ta so‘z fe‘l so‘z turkumiga, 7175 tasi ot so‘z turkumiga, 1189 ta sifat, 128 ta son, 390 ta olmosh, 344 ta ravish, 17 ta bog‘lovchi, 44 ta ko‘makchi, 15 ta yuklama, 105 ta modal so‘z, 81 ta undov so‘z, 14 ta taqlid so‘zlardan iborat (1-rasm).



1-rasm. Usmon Azim dramalarida so‘z turkumlarining taqsimlanishi.

Yuqorida keltirilgan diagrammadan shuni ko‘rish mumkinki, dramalar matnida statistik dominant – ot so‘z turkumi. Keyingi pog‘onada esa fe‘l so‘z turkumi. Shuningdek, til korpuslari yordamida uch xil statistik lug‘at tuzish mumkin:

Alfavit - chastotali lug‘at;

Chastotali lug‘at;

Ters lug‘at.

Xulosa qilib aytganda , Chastotali tahlil natijalari tilshunoslikda muhim rol o‘ynaydi. U tilning tuzilishini, grammatik va sintaktik qoidalarini turli davrlar nuqtayi nazaridan tadqiq qilishga yordam beradi, tilni o‘rganish jarayonini samarali qiladi. Korpus til hodisalarini har tomonlama, chuqur, ob`ektiv tadqiq etishga, lug‘at, grammatikalar yaratishga asos bo‘ladigan tizim. U tilni o‘rganishda, lingvodidaktika va lug‘at tuzishda mukammal hamda samaralidir.

ADABIYOTLAR

1. Мухаммедов С.М. Статистический анализ лексико-морфологической структуры узбекских газетных текстов: Автореф. дисс... канд. фил. наук.—Ташкент, 1980.;

2. Бабанаров А. Разработка принципов построения словарного обеспечения турецко-русского машинного перевода: Автореф. дисс... канд. фил. наук. — Ленинград, 1981.;

3. Айымбетов Н.К. Опыт лингвостатистического анализа лексики и морфологии каракалпакского публицистического текста: Автореф. дисс... канд. фил. наук.— Ташкент, 1987.;

4. Каримов С., Қаршиев А., Исроилова Г. Абдулла Қаҳҳор асарлари тилининг луғати. Алфавитли луғат. Частотали луғат. Терс луғат. — Тошкент, 2007. — 420 б.;

5. Ризаев С. Ўзбек тилининг лингвостатистик тадқиқи: Фил.фан.док.дис...автореф. — Тошкент, 2008.;

6. Мухаммедова С. Ҳаракат феъллари асосида компьютер дастурлари учун лингвистик таъмин яратиш. Методик қўлланма. — Тошкент, 2006. — 45 б.

7. Хамроева Ш.М. Ўзбек тили муаллифлик корпусини тузишнинг лингвистик асослари: Фил.фан.бўйича фалсафа доктори (PhD)...дис. — Бухоро, 2018. — 250 б.;

8. Тоирова Г.И. Ўзбек тили миллий корпусини яратишнинг назарий ва амалий масалалари. Филология фанлари доктори (DSc) диссертацияси автореферати. — Бухоро, 2021.

9. Ахмедова Д. Атов бирликларини ўзбек тили корпуслари учун лексик-семантик теглашнинг лингвистик асос ва моделлари Фил.фан.бўйича фалсафа доктори (PhD) дис. — Бухоро, 2020. — 40 б.;

10. Эшмуминов А.А. Ўзбек тили миллий корпуси синоним сўзлар базасини яратиш асослари: Фил.фан.бўйича фалсафа доктори (PhD) дис. — Термиз, 2019. — 140 б.

11. M.Yo‘ldoshev, S.Muhammedova, M.Saparniyozova “Matn lingvistikasi” Toshkent «Ishonchli hamkor» 2021